

Method and encoder for bit-rate saving encoding of audio signals

FIELD OF THE INVENTION

5 The invention relates to a method and an encoder for bit-rate saving encoding of audio signals, especially for encoding of audio signals according to MPEG 1 Audio Layer II.

10

BACKGROUND OF THE INVENTION

The MPEG 1 Audio standard as specified in ISO/IEC 11172-3 defines three operational modes known as Layers I, II and
15 III. Each Layer offers increased compression but also increased encoding complexity, whereby downward compatibility is guaranteed. Therefore, a Layer II decoder can also read a Layer I data stream but no Layer III data stream. Also a Layer III decoder can decode all MPEG 1
20 Audio bit streams, i. e. Layer I to III.

MPEG 1 Audio data compression is based on subband coding. The audio signal is split into 32 subbands of equal width. A quantization is performed using a psychoacoustic model
25 that is adapted to the masking behaviour of the human hearing. Each subband signal is quantized in such a way that the quantization noise introduced by the coding will not exceed the masking curve for that subband. After quantization the samples build - together with the scale
30 factors and further coding information - a frame structure for transmission.

In ISO/IEC 11172-3 two independent psychoacoustic models are defined which can be adapted to any Layer. The output

00629109.073100

from these psychoacoustic models is a set of Signal-to-Masking Ratios, SMR_n , for every subband n .

In order to calculate the SMR_n for the psychoacoustic
5 model 2 a Fast Fourier Transformation (FFT) with a length
of 1024 samples is used, which has to run two times per
channel, i. e. four times for a stereo channel.

10 SUMMARY OF THE INVENTION

The invention is based on the object of specifying a
method for bit-rate saving encoding of audio signals using
a psychoacoustical model with reduced computing power.
15 This object is achieved by means of the method specified
in claim 1.

It is a further object of the invention to disclose an
encoder which utilises the inventive method. This object
20 is achieved by the apparatus disclosed in claim 6.

The invention is based upon the following realization. On
the one hand a FFT is a special discrete Fourier
Transformation for which the number of samples has to be a
25 power of two, e.g. 1024 samples. However, on the other
hand the frame length of MPEG 1 Audio Layer II is 1152
samples which is no power of two. This results in the two
runs per channel of the FFT according to prior art.

30 The known formula for a discrete Fourier Transformation,
which calculates for L elements of a data series $z(n)$ the
corresponding L frequency values $F(m)$,

$$F(m) = \frac{1}{L} \sum_{n=0}^{L-1} z(n) \exp(-j2\pi \frac{mn}{L})$$

can be transformed first into k partial sums with M summands each

$$F(m) = \frac{1}{L} \sum_{i=0}^{k-1} \left[\sum_{n=0}^{M-1} z(kn+i) \exp(-j2\pi \frac{m(kn+i)}{L}) \right]$$

5

and finally with $L=kM$ and splitting the exponential function into

$$F(m) = \frac{1}{k} \sum_{i=0}^{k-1} \left[\frac{1}{M} \sum_{n=0}^{M-1} z(kn+i) \exp(-j2\pi \frac{mn}{L}) \right] \exp(-j2\pi \frac{mi}{L}).$$

10

Therefore, using k subtransformations with a length of $M=2^N$ allow time effective calculation with Fast Fourier Transformations, even if L is not a power of two.

15 In principle, the method for bit-rate saving encoding of audio signals using a psychoacoustic model, wherein a Fourier Transformation is performed for calculation of a minimum masking threshold and wherein L samples of the audio signal are arranged in a frame for transmission,
20 consists in the fact that the Fourier Transformation is performed with a length of L samples by calculating k subtransformations over 2^N samples with $k \cdot 2^N = L$ and fitting together the results of the k subtransformations.

25 In this way, the duplicate run per channel is avoided without lost of information or introduction of errors. Therefore, the computing power is reduced nearly to the half. This is especially important for the implementation on digital signal processors to run in real time.

30

Although the inventive method is especially advantageous, if the number k of subtransformations is not a power of 2,

090829109-073100

the use of the invention is not restricted to such values of k.

Advantageously, before fitting together the results of the k subtransformations, these are multiplied with phase correction factors.

In an advantageous manner, the Fourier Transformation is performed within the algorithm for the psychoacoustic model 2 of MPEG I Audio Layer II, wherein the frame length L is 1152 samples.

In an advantageous development k=9 subtransformations with a length of $M=2^N = 128$ samples are calculated.

BRIEF DESCRIPTION OF THE DRAWINGS

Exemplary embodiments of the invention are described with reference to the figure, which shows schematically the use of the subtransformations.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Although the encoder is not standardized in ISO/IEC 11172-3, certain means for encoding such as the estimation of the masking threshold or quantization are commonly used and therefore are not described in detail in the following. First of all PCM audio samples with a sampling frequency of 32, 44.1 or 48 kHz (or with half sample frequencies of 16, 22.05 or 24 kHz) are fed into the encoder. A mapping is performed to create a filtered and subsampled representation of the input audio stream in the form of so-called subband samples. Each subband signal is

quantized in such a way that the quantization noise introduced by the coding will not exceed the masking curve for that subband. In order to control the quantization a psychoacoustic model is used to calculate the new bit allocation, for MPEG 1 Audio Layer II for three blocks totaling 36 subband samples corresponding to 1152 input PCM samples. The calculation is based on the Signal-to-Masking Ratios, SMR_n , for all the subbands, which makes it necessary to determine for each subband the maximum signal level and the minimum masking threshold. The minimum masking threshold is derived by using the inventive method which is schematically shown in Fig. 1. After windowing the input PCM samples, samples corresponding to a frame F1 are split and fed to nine subtransformation ST1, ..., ST9. Each subtransformation has a length of 128 samples (9*128=1152) which is the seventh power of 2 ($2^7 = 128$). After Fast Fourier Transformations of the partial signals the results of the nine subtransformations are multiplied with phase correction factors, symbolized by multiplier units M1, ..., M9. The phase corrected data are combined, symbolized by adder unit A and then used for the further calculation of the psychoacoustic model. The same method is applied to the following audio samples corresponding to frames F2, F3 etc.

The invention can advantageously be used for encoding of audio signals according to MPEG 1 Audio Layer 2, but also applies to the encoding of any other kind of digital data.

The invention can be implemented in any kind of encoders which can be used e.g. for Digital Audio Broadcast (DAB), Cable and satellite radio/TV or digital recording devices like DVD-VRs etc.